

RESEARCH ARTICLE

# Genome-Wide Estimates of Coancestry, Inbreeding and Effective Population Size in the Spanish Holstein Population

Silvia Teresa Rodríguez-Ramilo<sup>1\*</sup>, Jesús Fernández<sup>1</sup>, Miguel Angel Toro<sup>2</sup>, Delfino Hernández<sup>3</sup>, Beatriz Villanueva<sup>1</sup>

**1** Departamento de Mejora Genética Animal, INIA, Madrid, Spain, **2** Departamento de Producción Animal, Escuela Técnica Superior de Ingenieros Agrónomos, Madrid, Spain, **3** Departamento Técnico Conafe, Madrid, Spain

\* [rodriguez.silvia@inia.es](mailto:rodriguez.silvia@inia.es)



## OPEN ACCESS

**Citation:** Rodríguez-Ramilo ST, Fernández J, Toro MA, Hernández D, Villanueva B (2015) Genome-Wide Estimates of Coancestry, Inbreeding and Effective Population Size in the Spanish Holstein Population. PLoS ONE 10(4): e0124157. doi:10.1371/journal.pone.0124157

**Academic Editor:** Peter James Hansen, University of Florida, UNITED STATES

**Received:** July 31, 2014

**Accepted:** March 10, 2015

**Published:** April 16, 2015

**Copyright:** © 2015 Rodríguez-Ramilo et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Data cannot be made publicly available due to legal restrictions. Requests to access data may be sent to the following: CONAFE contact: Sofia Alday [sofia.alday@conafe.com](mailto:sofia.alday@conafe.com) Managing Director CONAFE Apartado 31 28340 Valdemoro Spain, EUROGENOMICS contact: Clotilde Patry [clotilde.patry@uncea.fr](mailto:clotilde.patry@uncea.fr) UNCEIA-Maison Nationale des Eleveurs 149 rue de Bercy -75595 Paris Cedex 12 France.

**Funding:** This study was supported by the Ministerio de Economía y Competitividad, Spain (grant CGL2012-39861-C02-02).

## Abstract

Estimates of effective population size in the Holstein cattle breed have usually been low despite the large number of animals that constitute this breed. Effective population size is inversely related to the rates at which coancestry and inbreeding increase and these rates have been high as a consequence of intense and accurate selection. Traditionally, coancestry and inbreeding coefficients have been calculated from pedigree data. However, the development of genome-wide single nucleotide polymorphisms has increased the interest of calculating these coefficients from molecular data in order to improve their accuracy. In this study, genomic estimates of coancestry, inbreeding and effective population size were obtained in the Spanish Holstein population and then compared with pedigree-based estimates. A total of 11,135 animals genotyped with the Illumina BovineSNP50 BeadChip were available for the study. After applying filtering criteria, the final genomic dataset included 36,693 autosomal SNPs and 10,569 animals. Pedigree data from those genotyped animals included 31,203 animals. These individuals represented only the last five generations in order to homogenise the amount of pedigree information across animals. Genomic estimates of coancestry and inbreeding were obtained from identity by descent segments (coancestry) or runs of homozygosity (inbreeding). The results indicate that the percentage of variance of pedigree-based coancestry estimates explained by genomic coancestry estimates was higher than that for inbreeding. Estimates of effective population size obtained from genome-wide and pedigree information were consistent and ranged from about 66 to 79. These low values emphasize the need of controlling the rate of increase of coancestry and inbreeding in Holstein selection programmes.

## Introduction

It is well known that the effective size of a population ( $N_e$ ) is usually less than its census size as there are often deviations from the assumptions of the idealized population by having unequal

**Competing Interests:** The authors have declared that no competing interests exist.

sex ratios, variation in family size, unequal numbers in successive generations and overlapping generations [1]. A clear example of this is given by the Holstein cattle breed that despite being numerically very large (millions of animals spread across the world), shows a  $N_e$  of the order of 100 [2, 3] when calculated from the rate of inbreeding ( $\Delta F$ ) computed from pedigree data.

Holstein dairy cattle have dominated the milk production industry over decades. Intense and accurate artificial selection practised over many years has resulted in high rates of genetic gain for milk production traits. This has implied an extreme use of a limited number of elite sires of high genetic merit for production traits via artificial insemination. However, the high rates of gain have been accompanied by large increases in the rates at which inbreeding and coancestry ( $\Delta f$ ) accumulate and, consequently, by large reductions in  $N_e$ . This has led, in the last decades, to an increasing awareness of the need of measuring and controlling the loss of genetic variation and the increase in inbreeding to mitigate its negative effects [4]. These include inbreeding depression in production traits and reproductive ability [5, 6, 7, 8] and an increase in the prevalence of undesirable genetic disorders such as the complex vertebral malformation (CVM) [9], deficiency of uridine monophosphate synthase (DUMPS) [10], brachyspina syndrome (BS) [11] and Bovine leucocyte adhesion deficiency (BLAD) [12].

With the advances in high-throughput genotyping techniques and the development of chips containing thousands of SNPs at a reasonable cost, the implementation of genome-wide evaluation [13] has become routine in many large scale commercial Holstein breeding programmes [14, 15]. As high levels of accuracy can be achieved at an early age, generation intervals can be shortened leading to faster gain [16]. Although genomic selection leads to decreased rates of inbreeding per generation in comparison with BLUP selection [17, 18, 19], it is not inbreeding free. Thus, the need for measuring and controlling inbreeding remains [17].

Traditionally, coancestry and inbreeding coefficients have been estimated from pedigree records. However, the dense marker chips that have been developed for cattle with the main purpose of increasing selection responses, can be also used for obtaining genome-wide estimates of coancestry and inbreeding. Genomic estimates are expected to be more accurate than pedigree-based estimates because they i) reflect the actual percentage of the genome that is homozygous (inbreeding) or the actual percentage of the genome shared by two individuals (coancestry) whereas pedigree-based estimates are only expectations of such percentages; and ii) are able to capture relationships due to very distant common ancestors that pedigree-based estimates ignore [20].

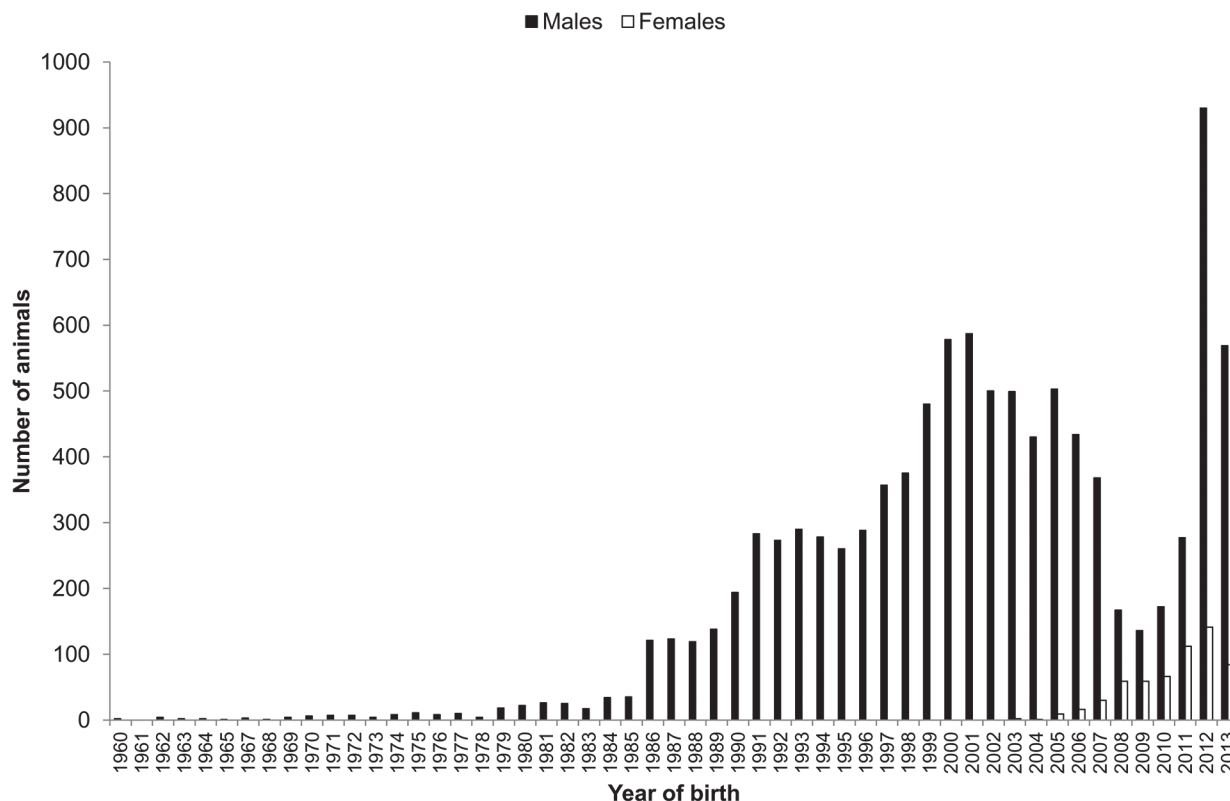
In livestock species, genomic estimates of inbreeding have been obtained for different cattle [8, 21, 22, 23, 24], sheep [25] and pig [26, 27] breeds, but estimates of genomic coancestry are rare [28]. However, both inbreeding and coancestry are of fundamental importance in animal breeding programmes. Although inbreeding depression depends on the levels of inbreeding and not on coancestry, in scenarios with non-random mating such as those in most cattle breeding programmes, the rate at which genetic variability is lost is better measured by  $\Delta f$  than by  $\Delta F$  [29].

The objective of this study was to obtain genomic estimates of coancestry and inbreeding coefficients and  $N_e$  in the Spanish Holstein population. Genomic estimates were then compared to those based on pedigree information.

## Materials and Methods

### Genomic and pedigree data

Genomic information from 11,135 animals belonging to the Spanish Holstein population was analyzed in this study. These individuals were genotyped with the Illumina BovineSNP50 BeadChip (versions v1 or v2). Only SNPs common to both chip versions were selected for the



**Fig 1. Distribution of genotyped animals by year of birth.**

doi:10.1371/journal.pone.0124157.g001

analysis (52,340 SNPs). SNPs positions within the genome were obtained from the UMD 3.0 bovine genome assembly [30]. Unmapped SNPs (523) and those mapped on chromosomes X or Y (1,056) were excluded. In addition, 14,068 SNPs with missing genotypes for more than 5% of the individuals were discarded. After that, 566 animals with more than 5% missing genotypes for the remaining 36,693 SNPs were also removed. The final dataset included 36,693 autosomal SNPs and 10,569 animals (9,990 bulls and 579 cows). The distribution of genotyped animals by year of birth is shown in Fig 1.

The pedigree data set, provided by the Spanish Holstein Association (Conafe), was constructed with all known ancestors of genotyped individuals and comprised 35,473 animals. The software Endog [31] was used to calculate the average numbers of generations traced (14.50), full traced generations (2.56) and equivalent complete generations (6.12). The generation interval ( $L$ ) was calculated as the average age of parents when their offspring was born. The average generation intervals for sires of bulls, dams of bulls, sires of cows and dams of cows were 6.06, 3.88, 5.24 and 3.77, respectively. The average weighted  $L$  across paths and years was 4.2 years. We observed that  $L$  has decreased across time. The average  $L$  for the periods 1960 to 1979, 1980 to 1999 and 2000 to 2013 were 5.1, 4.2 and 3.3, respectively.

## Genomic estimates of inbreeding and coancestry

Genomic estimates of inbreeding coefficients were obtained from runs of homozygosity (ROHs) which are long, uninterrupted stretches of homozygous genotypes [20]. Specifically, the inbreeding estimator,  $F_{ROH}$ , is the proportion of the genome that is in ROH. For individual

$i$ ,  $F_{ROH_i}$  was calculated as

$$F_{ROH_i} = \frac{\sum_{k=1}^{n_{ROH_i}} l_{ROH_{ik}}}{l_g},$$

where  $n_{ROH_i}$  is the total number of ROHs in individual  $i$ ,  $l_{ROH_{ik}}$  is the length of the  $k^{th}$  ROH in individual  $i$  in base pairs and  $l_g$  is the total length of the genome in base pairs.

The criteria used for defining a ROH were as follows: (i) the minimum length that constituted a ROH was 4 Mb; (ii) the minimum number of SNPs was 30; (iii) the minimum density was 1 SNP per 100 kb; (iv) the maximum distance allowed between two consecutive homozygous SNPs in a run was 1 Mb; and (v) a maximum of two missing genotypes and one heterozygous genotype within a particular ROH were permitted. The choice of 4 Mb for the minimum length permitted for defining a ROH was based on the results of Ferenčaković et al. [26] who showed that  $F_{ROH}$  based on shorter segments systematically overestimates autozygosity when using the Illumina BovineSNP50 BeadChip.

The coancestry coefficient based on IBD segments ( $f_{SEG}$ ) between individuals  $i$  and  $j$  was calculated as

$$f_{SEG_{ij}} = \frac{\sum_{k=1}^{n_{SEG_{ij}}} \sum_{a_i=1}^2 \sum_{b_j=1}^2 l_{SEG_k}(a_i, b_j)}{4l_g},$$

where  $l_{SEG_k}(a_i, b_j)$  is the length of the  $k^{th}$  shared IBD segment measured over homologue  $a$  of individual  $i$  and homologue  $b$  of individual  $j$  and  $n_{SEG_{ij}}$  is the total number of IBD segments between individuals  $i$  and  $j$ . The criteria for defining an IBD segment were equivalent to those used for defining a ROH. Both  $F_{ROH_i}$  and  $f_{SEG_{ij}}$  were obtained using the software developed by de Cara et al. [32].

In order to calculate coancestry coefficients based on IBD segments, haplotypes need to be inferred from genotype data. The phase inference was performed using the software BEAGLE [33] that uses localized haplotype clustering and fits the data through an Expectation-Maximization (EM) algorithm. Each chromosome was phased independently using the default parameters (10 iterations). After haplotype inference,  $f_{SEG_{ij}}$  was obtained as explained above.

## Genomic estimates of homozygosity and similarity

We also obtained genomic estimates of molecular homozygosity and molecular similarity on a SNP-by-SNP basis. The genomic similarity between individuals was obtained by applying Malécot's definition [34] to the marker loci; that is, the molecular similarity between individuals  $i$  and  $j$  ( $S_{SNP_{ij}}$ ) is the probability that two alleles at a given locus taken at random from each individual are equal (IBS). Analogously, the genomic homozygosity of individual  $i$  ( $H_{SNP_i}$ ) is the probability that the two alleles carried by this individual at a given locus are IBS. In this study,  $S_{SNP_{ij}}$  was calculated as

$$S_{SNP_{ij}} = (1/n_s) \sum_{s=1}^{n_s} \left[ \left( \sum_{k=1}^2 \sum_{m=1}^2 I_{s_{kjm_j}} \right) / 4 \right],$$

where  $n_s$  is the number of SNPs and  $I_{s_{kjm_j}}$  is the identity of the  $k^{th}$  allele from individual  $i$  with the  $m^{th}$  allele from individual  $j$  at SNP  $s$ , and takes a value of 1 if both alleles are identical and 0 otherwise. The genomic homozygosity of individual  $i$  was calculated as  $H_{SNP_i} = 2S_{SNP_{ii}} - 1$ , and was equal to the proportion of homozygous SNPs.

## Pedigree-based estimates

The pedigree-based coancestry ( $f_{PED_{ij}}$ ) and inbreeding ( $F_{PED_i}$ ) coefficients for the genotyped individuals were calculated by going back only five generations in the pedigree in order to homogenise the amount of pedigree information across individuals. This data set included 31,203 animals. Estimates of both coefficients were obtained using the software PEDIG [35] with the option that implements the algorithm of Meuwissen and Luo [36].

## Rates of change in coancestry and inbreeding, and effective population size

Rates of change in genomic and pedigree-based inbreeding per year ( $\Delta F_{ROH(y)}$  and  $\Delta F_{PED(y)}$ , respectively) were computed by regressing the natural logarithm of  $(1-F)$  for each individual on the year of birth. The slopes of these regressions are approximately equal to  $-\Delta F_{ROH(y)}$  and  $-\Delta F_{PED(y)}$ . Rates of change in inbreeding per generation ( $\Delta F_{ROH}$  and  $\Delta F_{PED}$ ) were calculated by multiplying the rates per year by the generation interval. Finally, estimates of  $N_e$  were obtained from the rate of change in inbreeding per generation as  $N_{eF\_ROH} = 1/2\Delta F_{ROH}$  and  $N_{eF\_PED} = 1/2\Delta F_{PED}$ . Rates of change in coancestry per year ( $\Delta f_{SEG(y)}$  and  $\Delta f_{PED(y)}$ ), and per generation ( $\Delta f_{SEG}$  and  $\Delta f_{PED}$ ), and effective population size ( $N_{ef\_SEG} = 1/2\Delta f_{SEG}$  and  $N_{ef\_PED} = 1/2\Delta f_{PED}$ ) were also computed following the same approach as for inbreeding. Rates of change in genomic homozygosity ( $\Delta H_{SNP(y)}$  and  $\Delta H_{SNP}$ ) and similarity ( $\Delta S_{SNP(y)}$  and  $\Delta S_{SNP}$ ), and the corresponding estimates of  $N_e$  ( $N_{eH\_SNP} = 1/2\Delta H_{SNP}$ , and  $N_{eS\_SNP} = 1/2\Delta S_{SNP}$ ) were also obtained.

## Results

Table 1 shows the mean, range, variance and coefficient of variation for the genomic homozygosity, similarity, inbreeding and coancestry coefficients. As expected, the molecular homozygosity and similarity coefficients were much higher than the inbreeding and coancestry coefficients. The reason is that the former (that are obtained on a SNP-by-SNP basis), do not discriminate alleles that are IBD or IBS. Estimates of  $F$  and  $f$ , based on ROHs and IBD segments, respectively, were close to the pedigree-based coefficients, indicating that they reflect well IBD. The highest coefficients of variation were observed for pedigree-based estimates.

Table 2 shows the relationships between the genomic homozygosity and similarity, and different inbreeding and coancestry coefficients. The correlation between  $F_{ROH}$  and  $F_{PED}$  was higher than that between  $H_{SNP}$  and  $F_{PED}$  but still relatively low. Actually, 33% and 26% of the variance in  $F_{PED}$  can be explained by  $F_{ROH}$  and  $H_{SNP}$ , respectively. The highest correlation was

**Table 1. Mean, range (minimum and maximum values), variance and coefficient of variation (CV) for different estimates.**

	Mean	Range	Variance	CV
$F_{ROH}$	0.0770	0.0000–0.2660	0.0010	0.4125
$F_{PED}$	0.0422	0.0000–0.2786	0.0007	0.6316
$H_{SNP}$	0.6451	0.4890–0.8769	0.0002	0.0206
$f_{SEG}$	0.0780	0.0000–0.5712	0.0006	0.3106
$f_{PED}$	0.0439	0.0000–0.4180	0.0005	0.5173
$S_{SNP}$	0.6447	0.5474–0.8366	0.0001	0.0148

$F_{ROH}$ : inbreeding based on ROHs;  $F_{PED}$ : pedigree-based inbreeding;  $H_{SNP}$ : SNP-by-SNP-based homozygosity;  $f_{SEG}$ : coancestry based on IBD segments;  $f_{PED}$ : pedigree-based coancestry;  $S_{SNP}$ : SNP-by-SNP-based similarity

doi:10.1371/journal.pone.0124157.t001

**Table 2. Intercept (*a*), regression coefficient (*b*) and correlation (*R*) between different estimates.**

Regression of			<i>a</i>	<i>b</i>	<i>R</i>
$F_{PED}$	on	$F_{ROH}$	0.01	0.48	0.57
$F_{PED}$	on	$H_{SNP}$	-0.62	1.03	0.51
$F_{ROH}$	on	$H_{SNP}$	-1.29	2.12	0.88
$f_{PED}$	on	$f_{SEG}$	-0.01	0.73	0.78
$f_{PED}$	on	$S_{SNP}$	-1.08	1.74	0.73
$f_{SEG}$	on	$S_{SNP}$	-1.49	2.43	0.95

$F_{PED}$ : pedigree-based inbreeding;  $F_{ROH}$ : inbreeding based on ROHs;  $H_{SNP}$ : SNP-by-SNP based homozygosity;  $f_{PED}$ : pedigree-based coancestry;  $f_{SEG}$ : coancestry based on IBD segments;  $S_{SNP}$ : SNP-by-SNP based similarity

doi:10.1371/journal.pone.0124157.t002

that between  $H_{SNP}$  and  $F_{ROH}$ . Seventy seven percent of the variance in  $F_{ROH}$  can be explained by  $H_{SNP}$ . Pearson's correlation coefficients involving  $S_{SNP}$ ,  $f_{SEG}$  and  $f_{PED}$  were substantially higher than those involving  $H_{SNP}$ ,  $F_{ROH}$ ,  $F_{PED}$ . In fact, 61% and 53% of the variance in  $f_{PED}$  is explained by  $f_{SEG}$  and  $S_{SNP}$ , respectively. Also, 90% of the variance in  $f_{SEG}$  is explained by  $S_{SNP}$ .

Table 3 shows the rates of change in genomic homozygosity and similarity and in genomic and pedigree-based inbreeding and coancestry per year and per generation, and the  $N_e$  estimated from those rates. Despite the different scales of the SNP-by-SNP based measures when compared with the inbreeding and coancestry coefficients (see Table 1), the rates of change were very similar for all these parameters, as expected. Consequently, estimates of  $N_e$  obtained from  $\Delta S_{SNP}$  and from  $\Delta f_{SEG}$  and  $\Delta f_{PED}$  were very close. Rates of change in  $H_{SNP}$ ,  $F_{SEG}$  and  $F_{PED}$  were also very close. Pedigree-based inbreeding rates were slightly lower than the corresponding coancestry rates. Similarly, rates of genomic homozygosity were slightly lower than the corresponding rates of genomic similarity. Consequently, estimates of  $N_e$  obtained from rates of change in molecular homozygosity and inbreeding were slightly higher than estimates obtained from rates of change in molecular similarity and coancestry.

Given that  $L$  decreased over time, we also estimated  $N_e$  for different periods (1960 to 1979, 1980 to 1999 and 2000 to 2013). Annual rates of homozygosity and inbreeding remained constant across time, but  $N_e$  increased as a consequence of the reduction in  $L$ . These estimates of  $N_e$  ranged from 61.3 to 65.4 for the period from 1960 to 1979, from 74.4 to 79.4 for the period from 1980 to 1999, and from 94.7 to 101.0 for the period from 2000 to 2013.

**Table 3. Rates of change in inbreeding, molecular homozygosity, coancestry and molecular similarity per year ( $\Delta F_{(y)}$ ,  $\Delta H_{(y)}$ ,  $\Delta f_{(y)}$  and  $\Delta S_{(y)}$ ), respectively, and per generation ( $\Delta F$ ,  $\Delta H$ ,  $\Delta f$  and  $\Delta S$ ) using different sources of information, and estimates of effective population sizes obtained from  $\Delta F$  ( $N_{eF}$ ),  $\Delta H$  ( $N_{eH}$ ),  $\Delta f$  ( $N_{ef}$ ) and from  $\Delta S$  ( $N_{eS}$ ).**

	ROHs or IBD segments	SNP-by-SNP	Pedigree
$\Delta F_{(y)}$ or $\Delta H_{(y)}$	0.0016	0.0016	0.0015
$\Delta F$ or $\Delta H$	0.0067	0.0067	0.0063
$N_{eF}$ or $N_{eH}$	74.4	74.4	79.4
$\Delta f_{(y)}$ or $\Delta S_{(y)}$	0.0016	0.0017	0.0018
$\Delta f$ or $\Delta S$	0.0067	0.0071	0.0076
$N_{ef}$ or $N_{eS}$	74.4	70.0	66.1

doi:10.1371/journal.pone.0124157.t003



## Discussion

In this study, estimates of coancestry, inbreeding and  $N_e$  obtained from genome-wide information have been compared with those obtained from pedigree information in the Spanish population of Holstein cattle. Our results confirm the small  $N_e$  of the Holstein breed and the need of controlling coancestry and inbreeding rates.

The correlation between  $F_{ROH}$  and  $F_{PED}$  (0.57) obtained in the present study is in the range of values obtained in previous studies. Ferenčaković et al. [22] found values ranging from 0.50 to 0.72 when analysing four different cattle breeds with the 50 K chip. Keller et al. [20] showed that  $F_{ROH}$  is preferable to  $F_{PED}$  and other measures of genomic inbreeding because it correlates better with inbreeding depression. More specifically, for a  $N_e$  similar to that estimated here for the Holstein population, they predicted that the correlation between homozygous mutation load and inbreeding coefficient was higher for  $F_{ROH}$  (0.60) than for  $F_{PED}$  (0.25).

Molecular homozygosity and similarity coefficients calculated on a SNP-by-SNP basis were much higher than pedigree-based inbreeding and coancestry coefficients because the latter refer to a base population where no homozygosity exists. Thus, with the SNP-by-SNP based method alleles that are IBD and IBS can not be distinguished. Several approaches have been proposed to express these genomic coefficients in the same scale as pedigree-based coefficients [37] but they require knowing the base population allele frequencies. However, given that these frequencies are usually unknown, these methods are generally inaccurate [38, 39]. Another reason for correcting the raw molecular homozygosity and / or similarity is that in genome-wide evaluation methods, genomic relationship matrices are usually combined with pedigree-based relationship matrices because SNP genotypes are not available for many animals included in the evaluation procedure [21, 40, 41, 42]. Genomic matrices are usually corrected using the frequencies estimated in the present population through the formula of VanRaden [43, 44]. An alternative would be to use genomic matrices based on IBD segments rather than are in the same scale as pedigree-based matrices. This could represent an advantage over current approaches used in the context of genome-wide selection for predicting breeding values [45].

Here, estimates of molecular homozygosity, molecular similarity, genomic inbreeding and coancestry were obtained using genotypes contained in the Illumina BovineSNP50 BeadChip. A higher density cattle chip (i.e., the BovineHD BeadChip) containing 777,962 SNPs is also commercially available and used in dairy cattle genetic evaluations. Although in principle it is expected that increasing marker density would lead to an increase in the accuracy of genomic predictions, there are indications that this is not the case at least when  $N_e$  is low [46]. Low  $N_e$  is associated with high linkage disequilibrium and, thus, increasing marker density above a certain level may not improve significantly the accuracy of genomic evaluations [47]. In fact, small differences have been detected in genomic prediction when comparing the high density panel with the 50 K chip [48]. In a similar way, two different studies have showed that the use of the high density chip does not lead to higher accuracies when estimating inbreeding [23, 49]. For three different cattle breeds, Ferenčaković et al. [23] found similar correlations between  $F_{ROH}$  and  $F_{PED}$  when using the 50 K chip (estimates ranging from 0.62 to 0.77) and when using the high density cattle panel (estimates ranging from 0.61 to 0.75). Also, Purfield et al. [49] analysed nine different cattle breeds and indicated that the percentages of the variance in  $F_{PED}$  explained by  $F_{ROH}$  with the high density and 50 k chips were very similar (56% and 53%, respectively).

As expected, rates of change in molecular homozygosity, and genomic and pedigree-based inbreeding (and rates of change in molecular similarity, and genomic and pedigree-based coancestry) were very similar [26, 38]. Rates of change were slightly lower for molecular homozygosity and inbreeding than for molecular similarity and coancestry, and consequently,  $N_{eH}$  and

$N_{eF}$  estimates were slightly higher than  $N_{eS}$  and  $N_{eF}$  estimates. This result could be a reflection of non-random mating (i. e. avoidance of matings between relatives) if the level of non-randomness was not constant across generations [50].

For cattle, there is evidence that the levels of  $N_e$  were large (of the order of tens of thousands or more) following domestication [2, 3], but current  $N_e$  in some modern breeds are of the order of 100 [2, 3]. Estimates of  $N_e$  obtained in this study across different time periods are close to those previously published for other Holstein populations. Estimates of  $N_e$  obtained from pedigree data have ranged from 39 in the US population [51] to 115 in the Canadian population [52]. Estimates obtained from genome-wide data have ranged from around 80 in the US population [53] to 150 in the Australian Holstein cattle [54].

The small  $N_e$  found in Holstein cattle reflects the fact that breeding strategies followed in this breed have implied a very heavy use of few top sires and reinforces the need of controlling the rate at which coancestry and inbreeding increase in selection programmes [2]. Both selection and mating strategies have been proposed in the past for controlling the rates of coancestry and inbreeding. In particular, optimum contribution selection [55, 56] provides a useful tool to manage the rate of accumulation of inbreeding. When applying this selection tool in cattle, higher genetic gains are expected at a given rate of inbreeding [4]. Both objectives can be attained using programmes where selection is based on classical BLUP-EBVs and those based on genomic EBVs [19]. The application of the method in Holstein cattle would however require a coordinated policy on the use of selected candidates and on breeding objectives [4]. This may require the cooperation of breeders and artificial insemination organizations to ensure that the correct animals and their contributions are identified [57, 58]. Recently, Pryce et al. [28] and Sun et al. [59] have showed that mating strategies based on genomic information can reduce progeny inbreeding when compared with strategies based on pedigree information. Thus, dense genetic marker information is not only valuable for increasing accuracies of selection but also for controlling the rate at which coancestry and inbreeding increase in dairy cattle breeding programmes [19].

## Acknowledgments

The authors would like to thank A. Caballero and one anonymous referee for valuable comments which helped to improve the manuscript. We are also very grateful to the EuroGenomics Consortium and Conafe for providing genotypes and pedigree data used in this study.

## Author Contributions

Conceived and designed the experiments: BV JF. Performed the experiments: STRR. Analyzed the data: STRR. Contributed reagents/materials/analysis tools: DH. Wrote the paper: STRR JF MAT DH BV.

## References

1. Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics. 4th edition. UK: Longman, Harlow.
2. Brotherstone S, Goddard M (2005) Artificial selection and maintenance of genetic variance in the global dairy cow population. *Philos Trans R Soc Lond B Biol Sci* 360: 1479–1488. PMID: [16048790](#)
3. Hill WG (2010) Understanding and using quantitative genetic variation. *Philos Trans R Soc Lond B Biol Sci* 365: 73–85. doi: [10.1098/rstb.2009.0203](#) PMID: [20008387](#)
4. Kearney JF, Wall E, Villanueva B, Coffey MP (2004) Inbreeding trends and application of optimized selection in the UK Holstein population. *J Dairy Sci* 87: 3503–3509. PMID: [15377628](#)
5. González-Recio O, López de Maturana E, Gutiérrez JP (2007) Inbreeding depression on female fertility and calving ease in Spanish dairy cattle. *J Dairy Sci* 90: 5744–5752. PMID: [18024768](#)



6. Parland SM, Kearney JF, Rath M, Berry DP (2007) Inbreeding effects on milk production, calving performance, fertility and conformation in Irish Holstein-Friesians. *J Dairy Sci* 90: 4411–4419. PMID: [17699061](#)
7. Miglior F, Van Doormaal BJ, Kistemaker G (2008) Phenotypic analysis of inbreeding depression for traits measured in Canadian dairy cattle breeds. [http://www.cdn.ca/Articles/GEbmAY2008/Industry\\_Session\\_Agenda-May2008.pdf](http://www.cdn.ca/Articles/GEbmAY2008/Industry_Session_Agenda-May2008.pdf).
8. Bjelland DW, Weigel KA, Vukasinovic N, Nkrumah JD (2013) Evaluation of inbreeding depression in Holstein cattle using whole-genome SNP markers and alternative measures of genomic inbreeding. *J Dairy Sci* 96: 4697–4706. doi: [10.3168/jds.2012-6435](#) PMID: [23684028](#)
9. Schütz E, Scharfenstein M, Brenig B (2008) Implication of complex vertebral malformation and bovine leukocyte adhesion deficiency DNA-based testing on disease frequency in the Holstein population. *J Dairy Sci* 91: 4854–4859. doi: [10.3168/jds.2008-1154](#) PMID: [19038961](#)
10. Shanks RD, Dombrowski DB, Harpestad GW, Robinson JL (1984) Inheritance of UMP synthase in dairy cattle. *J Hered* 75: 337–340. PMID: [6548235](#)
11. Charlier C, Agerholm JS, Coppieters W, Karlsson-Mortensen P, Li W, de Jong G, et al. (2012) A deletion in the bovine *FANCI* gene compromises fertility by causing fetal death and brachyspina. *PLoS ONE* 7 (8): e43085. doi: [10.1371/journal.pone.0043085](#) PMID: [22952632](#)
12. Kehrli ME Jr, Schmalstieg FC, Anderson DC, Van der Maaten MJ, Hughes BJ, Ackermann MR, et al. (1990) Molecular definition of the bovine granulocytopenia syndrome: Identification of deficiency of the Mac-1 (CD11b/CD18) glycoprotein. *Am J Vet Res* 51: 1826–1836. PMID: [1978618](#)
13. Meuwissen TH, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157: 1819–1829. PMID: [11290733](#)
14. Lund M, de Ross APW, de Vries A, Druet T, Ducrocq V, Fritz S, et al. (2011) A common reference population from four European Holstein populations increases reliability of genomic predictions. *Genet Sel Evol* 43: 43. doi: [10.1186/1297-9686-43-43](#) PMID: [22152008](#)
15. VanRaden PM, Cooper T (2011) The genomic evaluation system in the United States: Past, present, future. *J Dairy Sci* 94: 3202–3211. doi: [10.3168/jds.2010-3866](#) PMID: [21605789](#)
16. Schaeffer LR (2006) Strategy for applying genome-wide selection in dairy cattle. *J Anim Breed Genet* 123: 218–223. PMID: [16882088](#)
17. Daetwyler HD, Villanueva B, Bijma P, Woolliams JA (2007) Inbreeding in genome-wide selection. *J Anim Breed Genet* 124: 369–376. PMID: [18076474](#)
18. de Roos APW, Schrooten C, Veerkamp RF, van Arendonk JAM (2011) Effects of genomic selection on genetic improvement, inbreeding, and merit of young versus proven bulls. *J Dairy Sci* 94: 1559–1567. doi: [10.3168/jds.2010-3354](#) PMID: [21338821](#)
19. Sonesson AK, Woolliams JA, Meuwissen TH (2012) Genomic selection requires genomic control of inbreeding. *Genet Sel Evol* 44: 27. doi: [10.1186/1297-9686-44-27](#) PMID: [22898324](#)
20. Keller MC, Visscher PM, Goddard ME (2011) Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. *Genetics* 189: 237–249. doi: [10.1534/genetics.111.130922](#) PMID: [21705750](#)
21. VanRaden PM, Olson KM, Wiggans GR, Cole JB, Tooker ME (2011) Genomic inbreeding and relationships among Holstein, Jersey, and Brown Swiss. *J Dairy Sci* 94: 5673–5682. doi: [10.3168/jds.2011-4500](#) PMID: [22032391](#)
22. Ferenčaković M, Hamzić E, Gredler B, Solberg TR, Klemetsdal G, Curik I, et al. (2012) Estimates of autozygosity derived from runs of homozygosity: empirical evidence from selected cattle populations. *J Anim Breed Genet* 130: 286–293. doi: [10.1111/jbg.12012](#) PMID: [23855630](#)
23. Ferenčaković M, Sölkner J, Curik I (2013) Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. *Genet Sel Evol* 45: 42. doi: [10.1186/1297-9686-45-42](#) PMID: [24168655](#)
24. Gómez-Romano F, Sölkner J, Villanueva B, Mészáros G, de Cara MAR, Pérez O'Brien AM, et al. (2014) Genomic estimates of inbreeding and coancestry in Brown Swiss cattle. Proceedings of the 10th World Congress of Genetics Applied to Livestock Production. 17–22 August, Vancouver, Canada.
25. Li M-H, Strandén I, Tiirikka T, Sevón-Aimonen M-L, Kantanen J (2011) A comparison of approaches to estimate the inbreeding coefficient and pairwise relatedness using genomic and pedigree data in a sheep population. *PLoS ONE* 6(11): e26256. doi: [10.1371/journal.pone.0026256](#) PMID: [22114661](#)
26. Saura M, Fernández A, Rodríguez MC, Toro MA, Barragán C, Fernández AI, et al. (2013) Genome-wide estimates of coancestry and inbreeding in a closed herd of ancient Iberian pigs. *PLoS ONE* 8(10): e78314. doi: [10.1371/journal.pone.0078314](#) PMID: [24205195](#)

27. Silió L, Rodríguez MC, Fernández A, Barragán C, Benítez R, Óvilo C, et al. (2013) Measuring inbreeding and inbreeding depression on pig growth from pedigree or SNP-derived metrics. *J Anim Breed Genet* 130: 349–360. doi: [10.1111/jbg.12031](https://doi.org/10.1111/jbg.12031) PMID: [24074172](https://pubmed.ncbi.nlm.nih.gov/24074172/)
28. Pryce JE, Hayes BJ, Goddard ME (2012) Novel strategies to minimize progeny inbreeding while maximizing genetic gain using genomic information. *J Dairy Sci* 95: 377–388. doi: [10.3168/jds.2011-4254](https://doi.org/10.3168/jds.2011-4254) PMID: [22192217](https://pubmed.ncbi.nlm.nih.gov/22192217/)
29. Toro MA, Villanueva B, Fernández J (2014) Genomics and conservation genetics. *Livest Sci* 166: 48–53.
30. Zimin AV, Delcher AL, Florea L, Kelley DR, Schatz MC, Puiu D, et al. (2009) A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biol* 10: R42. doi: [10.1186/gb-2009-10-4-r42](https://doi.org/10.1186/gb-2009-10-4-r42) PMID: [19393038](https://pubmed.ncbi.nlm.nih.gov/19393038/)
31. Gutiérrez JP, Goyache F (2005) A note on ENDOG: a computer program for analysing pedigree information. *J Anim Breed Genet* 122: 172–176. PMID: [16130468](https://pubmed.ncbi.nlm.nih.gov/16130468/)
32. de Cara MAR, Villanueva B, Toro MA, Fernández J (2013) Using genomic tools to maintain diversity and fitness in conservation programmes. *Mol Ecol* 22: 6091–6099. doi: [10.1111/mec.12560](https://doi.org/10.1111/mec.12560) PMID: [24128280](https://pubmed.ncbi.nlm.nih.gov/24128280/)
33. Browning SR, Browning BL (2007) Rapid and accurate haplotype phasing and missing data inference for whole genome association studies using localized haplotype clustering. *Am J Hum Genet* 81: 1084–1097. PMID: [17924348](https://pubmed.ncbi.nlm.nih.gov/17924348/)
34. Malécot G (1948) *Les mathématiques de l'hérédité*. Paris: Masson & Cie.
35. Boichard D (2002) PEDIG: A Fortran package for pedigree analysis suited for large populations. *Proc 7th World Cong Genet Appl Livest Prod*, Montpellier, France. CD-ROM communication no. 28–13.
36. Meuwissen TH, Luo Z (1992) Computing inbreeding coefficients in large populations. *Genet Sel Evol* 24: 305–313.
37. Toro MA, García-Cortés LA, Legarra A (2011) A note on the rationale for estimating genealogical coancestry from molecular markers. *Genet Sel Evol* 43: 27. doi: [10.1186/1297-9686-43-27](https://doi.org/10.1186/1297-9686-43-27) PMID: [21749687](https://pubmed.ncbi.nlm.nih.gov/21749687/)
38. Toro MA, Barragán C, Óvilo C, Rodrigáñez J, Rodríguez C, Silió L (2002) Estimation of co-ancestry in Iberian pigs using molecular markers. *Conserv Genet* 3: 309–320.
39. Rodríguez-Ramilo ST, Toro MA, Martínez P, Castro J, Bouza C, Fernández J (2007). Accuracy of pairwise methods in the reconstruction of family relationships, using molecular information from turbot (*Scophthalmus maximus*). *Aquaculture* 273: 434–442.
40. Forni S, Aguilar I, Misztal I (2011) Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. *Genet Sel Evol* 43: 1. doi: [10.1186/1297-9686-43-27](https://doi.org/10.1186/1297-9686-43-27) PMID: [21749687](https://pubmed.ncbi.nlm.nih.gov/21749687/)
41. Meuwissen TH, Luan T, Woolliams JA (2011) The unified approach to the use of genomic and pedigree information in genomic evaluations revisited. *J Anim Breed Genet* 128: 429–439. doi: [10.1111/j.1439-0388.2011.00966.x](https://doi.org/10.1111/j.1439-0388.2011.00966.x) PMID: [22059576](https://pubmed.ncbi.nlm.nih.gov/22059576/)
42. Rodríguez-Ramilo ST, García-Cortés LA, González-Recio Ó (2014) Combining genomic and genealogical information in a Reproducing Kernel Hilbert Spaces regression model for genome-enabled predictions in dairy cattle. *PLoS ONE* 9(3): e93424. doi: [10.1371/journal.pone.0093424](https://doi.org/10.1371/journal.pone.0093424) PMID: [24671175](https://pubmed.ncbi.nlm.nih.gov/24671175/)
43. VanRaden PM (2007) Genomic measures of relationship and inbreeding. *Interbull Bull* 37: 33–36.
44. VanRaden PM (2008) Efficient methods to compute genomic predictions. *J Dairy Sci* 91: 4414–4423. doi: [10.3168/jds.2007-0980](https://doi.org/10.3168/jds.2007-0980) PMID: [18946147](https://pubmed.ncbi.nlm.nih.gov/18946147/)
45. Luan T, Yu X, Dolezal M, Bagnato A, Meuwissen THE (2014) Genomic prediction based on runs of homozygosity. *Genet Sel Evol* 46: 64. PMID: [25284459](https://pubmed.ncbi.nlm.nih.gov/25284459/)
46. MacLeod IM, Hayes BJ, Goddard ME (2013) Will sequence SNP data improve the accuracy of genomic prediction in the presence of long term selection? *Proc Assoc Advmt Anim Breed Genet* 20: 215–219.
47. Gómez-Romano F, Villanueva B, de Cara MAR, Fernández J (2013) Maintaining genetic diversity using molecular coancestry: the effect of marker density and effective population size. *Genet Sel Evol* 45: 38. doi: [10.1186/1297-9686-45-38](https://doi.org/10.1186/1297-9686-45-38) PMID: [24088414](https://pubmed.ncbi.nlm.nih.gov/24088414/)
48. Su G, Brøndum RF, Ma P, Guldbrandsen B, Aamand GP, Lund MS (2012) Comparison of genomic predictions using medium-density (54000) and high-density (777000) single nucleotide polymorphism marker panels in Nordic Holstein and Red Dairy Cattle populations. *J Dairy Sci* 95: 4657–4665. doi: [10.3168/jds.2012-5379](https://doi.org/10.3168/jds.2012-5379) PMID: [22818480](https://pubmed.ncbi.nlm.nih.gov/22818480/)
49. Purfield DC, Berry DP, Mc Parland S, Bradley DG (2012) Runs of homozygosity and population history in cattle. *BMC Genet* 13: 70. doi: [10.1186/1471-2156-13-70](https://doi.org/10.1186/1471-2156-13-70) PMID: [22888858](https://pubmed.ncbi.nlm.nih.gov/22888858/)

50. Caballero A (1994) Developments in the prediction of effective population size. *Heredity* 73: 657–679. PMID: [7814264](#)
51. Weigel KA (2001) Controlling inbreeding in modern breeding programs. *J Dairy Sci* 84: E177–E184.
52. Stachowicz K, Sargolzaei M, Miglior F, Schenkel FS (2011) Rates of inbreeding and genetic diversity in Canadian Holstein and Jersey cattle. *J Dairy Sci* 94: 5160–5175. doi: [10.3168/jds.2010-3308](#) PMID: [21943766](#)
53. Sargolzaei M, Schenkel FS, Jansen GB, Schaeffer LR (2008) Extent of linkage disequilibrium in Holstein cattle in North America. *J Dairy Sci* 91: 2106–2117. doi: [10.3168/jds.2007-0553](#) PMID: [18420642](#)
54. Hayes BJ, Visscher PM, McPartlan HC, Goddard ME (2003) Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Res* 13: 635–643. PMID: [12654718](#)
55. Meuwissen TH (1997) Maximizing the response of selection with a predefined rate of inbreeding. *J Anim Sci* 75:934–940. PMID: [9110204](#)
56. Grundy B, Villanueva B, Woolliams JA (1998) Dynamic selection procedures for constrained inbreeding and their consequences for pedigree development. *Genet Res* 72:159–168.
57. Weigel KA, Lin SW (2002) Controlling inbreeding by constraining the average relationship between parents of young bulls entering AI progeny test programs. *J Dairy Sci* 85: 2376–2383. PMID: [12362471](#)
58. Avendaño SA, Villanueva B, Woolliams JA (2003) Expected increases in genetic merit from using optimized contributions in two livestock populations of beef cattle and sheep. *J Anim Sci* 81: 2964–2975. PMID: [14677851](#)
59. Sun C, VanRaden PM, O'Connell JR, Weigel KA, Gianola D (2013) Mating programs including genomic relationships and dominance effects. *J Dairy Sci* 96: 8014–8023. doi: [10.3168/jds.2013-6969](#) PMID: [24119810](#)